

A CHARACTERIZATION OF MAXIMUM INDEPENDENT SETS OF DE BRUIJN GRAPHS

DUSTIN A. CARTWRIGHT, MARÍA ANGÉLICA CUETO,
AND ENRIQUE A. TOBIS[†]

ABSTRACT. The de Bruijn graph $B(d, 3)$ consists of all strings of length 3, taken from an alphabet of size d , with edges between words which are distinct substrings of a word of length 4. We give an inductive characterization of the maximum independent sets of the de Bruijn graphs $B(d, 3)$ and for the de Bruijn graph with loops removed, for all d . We derive a recurrence relation for their number.

1. INTRODUCTION

For any positive integers d and D , the *de Bruijn* graph $B(d, D)$ is the directed graph whose d^D nodes consist of all the D -digit words from the alphabet $\{0, \dots, d-1\}$. There is an edge from a word $x = x_1 \dots x_D$ to $y = y_1 \dots y_D$ if and only if $x_2 \dots x_D = y_1 \dots y_{D-1}$. It is worth mentioning that $B(d, D+1)$ is the line graph of $B(d, D)$. These graphs were introduced in the paper “A Combinatorial Problem” by N. G. de Bruijn, under the name of *T-nets* [5]. Since then, de Bruijn graphs have been used in a wide range of contexts. For example, in the study of normal numbers [15, 1], network topology [2, 11], quantum computation [13], and sequence assembly [12].

The graph $B(d, D)$ contains d nodes of the form $xx \dots xx$, which have an edge to themselves. In a slight abuse of notation, we will refer to such a node as *the loop* x . Notice that a loop cannot be in any independent set of $B(d, D)$. We study the structure of the maximum independent sets of $B(d, 3)$. Since they cannot have loops, we call them *loop-less maximum independent sets* (LMISs) of $B(d, 3)$. The loop-less

2000 *Mathematics Subject Classification.* 05C69, 05A15.

Key words and phrases. de Bruijn graphs, maximum independent sets.

[†]Corresponding author.

D. A. Cartwright was supported by the National Science Foundation grants DMS-0354321, DMS-0456960, and DMS-0757236. M. A. Cueto was supported by a UC Berkeley Chancellor’s Fellowship. E. A. Tobis was supported by a CONICET doctoral fellowship, CONICET PIP 5617, ANPCyT PICT 20569 and UBACyT X042 and X064 grants.

maximum independent sets of $B(d, 3)$ are precisely the comma-free codes of length 3, introduced in [4] in an attempt to explain the genetic code, and later generalized in [6, 8, 7]. Following the literature, we also study the maximum independent sets of the graph $B(d, 3)$ with the edges $xx \dots xx \rightarrow xx \dots xx$ removed. We call them *maximum independent sets* (MISs) of $B(d, 3)$. As we will see, they all contain loops. Following [10], we write $\alpha(d, D)$ for the size of an MIS with loops and $\alpha^*(d, D)$ for the size of a loop-less MIS.

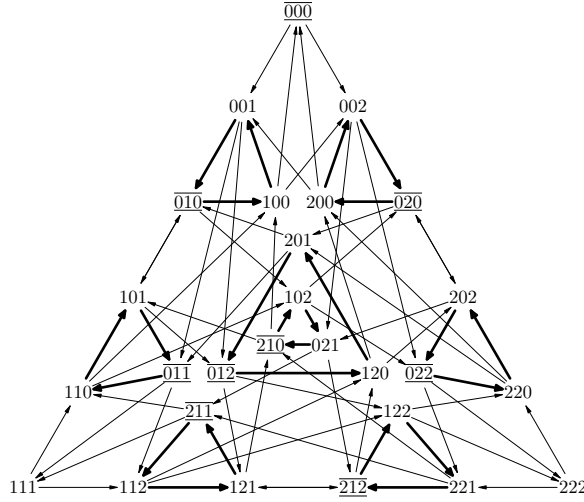


FIGURE 1. The nodes of an MIS of $B(3, 3)$ are highlighted. Bold arrows indicate edges under the shift function θ . The loops on 000, 111, and 222 are not shown.

The maximum independent sets of de Bruijn graphs have been previously studied ([9] and [10], for example). From [10], one can show that for $d \geq 4$, $\alpha(d, 2) = \alpha^*(d, 2) = \lfloor d^2/4 \rfloor$ and the number of maximum independent sets of $B(d, 2)$ is $\binom{d}{d/2}$ if d is even and $2\binom{d}{(d-1)/2}$ if d is odd. In the present work, we study the MISs of the graphs $B(d, 3)$. Again, in [10], it was proved that

$$(1) \quad \alpha(d, 3) = \frac{d^3 - d}{3} + 1 \quad \text{and} \quad \alpha^*(d, 3) = \frac{d^3 - d}{3}.$$

We define four functions which, together with the action of the symmetric group S_d recursively generate all MISs in $B(d, 3)$. From this, we deduce a recurrence relation for the number of MISs in $B(d, 3)$.

From Theorem 4.5 in Section 4, our main result, we derive the following:

Theorem: If we let a_d be the number of maximum independent sets of $B(d, 3)$, then a_d has exponential generating function

$$\sum_{d=1}^{\infty} \frac{a_d t^d}{d!} = \frac{t + t^2}{1 - 2t - t^2}.$$

We obtained the initial data for this work by computing the first few terms of a_d using the free computer algebra system CoCoA ([3]). We have computed all the MISs for $d \leq 5$ by an exhaustive procedure, and we used those sets in some of our proofs.

The rest of this paper is organized as follows. In Section 2, we define two functions f and f' that take a maximum independent set of $B(d, 3)$ and return maximum independent sets of $B(d + 1, 3)$. We construct another two functions g and g' that take a maximum independent set of $B(d, 3)$ and return a maximum independent set of $B(d + 2, 3)$. We also prove some basic facts about the images of these four functions.

In Section 3 we present the interactions between the symmetric group and the functions defined in Section 2. We compute the stabilizers of the images of our four functions. We later prove that all four functions map orbits under the action of the symmetric group into orbits under that same action. Furthermore, we show that the images of all four functions extended to orbits are disjoint. These results are used to prove the main theorems of this paper in Section 4.

In Section 5 we show the structure of the loop-less maximum independent sets of $B(d, 3)$, and show that their number coincides with the number of maximum independent sets of $B(d, 3)$ with loops.

2. INDUCTIVE CONSTRUCTION OF MAXIMUM INDEPENDENT SETS

In this section, we present four combinatorial operations that transform a maximum independent set in the de Bruijn graph $B(d, 3)$ into a maximum independent set in either $B(d + 1, 3)$ or $B(d + 2, 3)$.

Essential to the structure of the de Bruijn graph $B(d, 3)$ are the loops and the θ -cycles. Lichiardopol defined a shift function θ on the nodes of $B(d, 3)$ by $\theta(xyz) = yzx$. If xyz is not a loop, then xyz , $\theta(xyz)$, and $\theta^2(xyz)$ form a directed 3-cycle. In an abuse of notation, we will sometimes use θ to mark the edges of this 3-cycle. The action of θ induces a decomposition of $B(d, 3)$ into $(d^3 - d)/3$ cycles of length 3, and d cycles of length 1 (i.e. the loops.) Each of these disjoint cycles contributes at most one node to any independent set of $B(d, 3)$. Whenever we speak of “cycles”, we mean the cycles induced by θ . In Figure 1, the θ -cycles are indicated by darker cycles.

Convention 2.1. Throughout this paper, we let $[d]$ stand for the set $\{0, \dots, d-1\}$.

Definition 2.2. Let A be a set of nodes from $B(d, 3)$. Let x and y be two digits in $[d]$. We say that y *appears between x in A* if the node $xyx \in A$. We define $\mathcal{M}_x(A)$ as the set of digits which do not appear between x in A . We define $m_x(A)$ as the number of digits which do not appear between x in A , i.e. $m_x(A) = |\mathcal{M}_x(A)|$.

Proposition 2.3. *A maximum independent set S of $B(d, 3)$ contains either one or two loops. S always contains one loop a such that $m_a(S) = 0$. If S contains another loop b , then $m_b(S) = 1$ and $\mathcal{M}_b(S) = \{a\}$. In particular, S contains aba , but not bab .*

Proof. Suppose that an independent set S contains two loops a and b . Consider the two cycles $aab \xrightarrow{\theta} aba \xrightarrow{\theta} baa$ and $bba \xrightarrow{\theta} bab \xrightarrow{\theta} abb$. These two cycles can contribute at most the nodes aba and bab , since aab, baa, bba and abb are adjacent to either aaa or bbb . However, aba and bab are adjacent to each other, so only one of them can be present in S . Therefore, for every pair of loops present in an independent set S , one of the $(d^3 - d)/3$ cycles contributes no nodes to S .

If l is the number of loops in S , there are $\binom{l}{2}$ pairs of loops, and thus $\binom{l}{2}$ cycles which cannot contribute any nodes to S . Therefore, $|S| \leq (d^3 - d)/3 + l - \binom{l}{2}$, so for $l > 2$, S cannot be a maximum independent set.

The rest of the proposition follows from the previous observations: S must have one loop, by cardinality. If S is a maximum independent set with only one loop a , then every cycle contributes one node. The cycles $aax \xrightarrow{\theta} axa \xrightarrow{\theta} xaa$ can only contribute axa . If S has two loops a and b , then only one of aba and bab can be present. \square

Remark 2.4. The previous proof illustrates a tool we use throughout this paper. When dealing with an MIS S , we often choose a cycle $xyz \xrightarrow{\theta} yzx \xrightarrow{\theta} zxy$ and show that two of its nodes cannot be in S , therefore concluding that the third one is in S . Sometimes, we show that none of the three are in S , thus contradicting the fact that S is maximum.

Convention 2.5. From now on, we state things like “Let S be a maximum independent set with loops a and possibly b ,” when in fact S may have just one loop. In that case, everything said about b should be disregarded.

Definition 2.6. If w is a node, we will denote by $w[x \rightarrow y]$ the node that results from replacing every occurrence of the digit x by the digit

y in w . We write $x \in w$ to mean that x is one of the digits that appear in w .

We denote by a the loop of S such that $m_a(S) = 0$. If S has another loop we denote it b .

Definition 2.7. Let S be a maximum independent set of $B(d, 3)$. We define $f(S) \subset B(d+1, 3)$ to be the union of S with the sets

$$\begin{aligned} U_1(S) &= \{w[a \rightarrow d] \mid w \in S, a \in w, w \neq aaa, w \neq aba\}, \\ U_2(S) &= \{axd \mid x \in [\mathbf{d}] \setminus \{a, b\}\}, \\ U_3(S) &= \{dxa \mid x \in [\mathbf{d}] \setminus \{a, b\}\}, \\ U_4(S) &= \{udv \mid u, v \in \{a, b\}\}, \\ U_5(S) &= \{udd \mid u \in \{a, b\}\}. \end{aligned}$$

Remark 2.8. The loops of S are the same as the loops of $f(S)$.

Proposition 2.9. *If S is a maximum independent set of $B(d, 3)$, then $f(S)$ is a maximum independent set of $B(d+1, 3)$.*

Proof. $f(S)$ is made up of six disjoint sets. Let l be the number of loops of S . We have

$$\begin{aligned} |S| &= \frac{d^3 - d}{3} + 1, \\ |U_1(S)| &= (d-1)(d-1) - (l-1) + (d-l) = d^2 - d + 2 - 2l \\ |U_2(S)| &= |U_3(S)| = (d-l), \\ |U_4(S)| &= l^2, \\ |U_5(S)| &= l. \end{aligned}$$

Only the cardinality of $U_1(S)$ requires explanation. Notice that $B(d, 3)$ has $(d-1)(d-1)$ cycles which contain a once, with the exception of $abb \xrightarrow{\theta} bba \xrightarrow{\theta} bab$ in the case that $l = 2$. Each of these contributes one element to S and thus to $U_1(S)$. In addition, S contains one element from each of the $d-l$ cycles of the form $aa\bar{x} \xrightarrow{\theta} a\bar{x}a \xrightarrow{\theta} \bar{x}aa$, where \bar{x} is not a loop.

We add the six quantities to obtain

$$|f(S)| = \frac{(d+1)^3 - (d+1)}{3} + (l-1)(l-2) + 1.$$

Since l is either 1 or 2, $f(S)$ has the size of an MIS.

We still have to prove that $f(S)$ is an independent set. This amounts to noticing that there are no arrows between the six sets defining $f(S)$.

The only remark to bear in mind is that axa is in S for all x , and that $bx b$ is also in S , except for $x = a$. We leave the details to the reader. \square

We define another function very similar to f .

Definition 2.10. Let S be a maximum independent set of $B(d, 3)$. We define $f'(S) \subset B(d+1, 3)$ as the union of S , the sets $U_1(S)$, $U_2(S)$, $U_3(S)$, $U_4(S)$ from Definition 2.7, and

$$U'_5(S) = \{ddu \mid u \in \{a, b\}\}.$$

Proposition 2.11. *If S is a maximum independent set of $B(d, 3)$, then $f'(S)$ is a maximum independent set of $B(d+1, 3)$.*

Proof. This proposition is proved analogously to Proposition 2.9. \square

Definition 2.12. Let S be a maximum independent set of $B(d, 3)$. We define $g(S) \subset B(d+2, 3)$ to be the union of S with the sets

$$\begin{aligned} V_1(S) &= \{w[a \rightarrow y] \mid y \in \{d, d+1\}, w \in S, a \in w, w \neq aaa, w \neq aba\}, \\ V_2(S) &= \{axy \mid x \in [\mathbf{d}] \setminus \{a, b\}, y \in \{d, d+1\}\}, \\ V_3(S) &= \{yxa \mid x \in [\mathbf{d}] \setminus \{a, b\}, y \in \{d, d+1\}\}, \\ V_4(S) &= \{yxz \mid y, z \in \{d, d+1\}, y \neq z, x \in [\mathbf{d}] \setminus \{a, b\}\}, \\ V_5(S) &= \{uyv \mid u, v \in \{a, b\}, y \in \{d, d+1\}\}, \\ V_6(S) &= \{uyy \mid u \in \{a, b\}, y \in \{d, d+1\}\}, \\ V_7(S) &= \{yzu \mid y, z \in \{d, d+1\}, y \neq z, u \in \{a, b\}\}, \\ V_8(S) &= \{d(d+1)(d+1), (d+1)dd\}. \end{aligned}$$

Proposition 2.13. *If S is a maximum independent set of $B(d, 3)$, then $g(S)$ is a maximum independent set of $B(d+2, 3)$.*

Proof. $g(S)$ is made up of nine disjoint sets. For each pair of sets, it is clear that there are no edges between them. We now show that $g(S)$

has the right size. If l is the number of loops of S ,

$$\begin{aligned} |S| &= \frac{d^3 - d}{3} + 1, \\ |V_1(S)| &= 2|U_1(S)| = 2(d^2 - d + 2 - 2l), \\ |V_2(S)| &= 2|U_2(S)| = 2(d - l), \\ |V_3(S)| &= 2|U_3(S)| = 2(d - l), \\ |V_4(S)| &= 2|U_4(S)| = 2l^2, \\ |V_5(S)| &= 2(d - l), \\ |V_6(S)| &= 2|U_5(S)| = 2l, \\ |V_7(S)| &= 2l, \\ |V_8(S)| &= 2. \end{aligned}$$

The sum of these sizes is

$$|g(S)| = \frac{(d+2)^3 - (d+2)}{3} + 2(l-1)(l-2) + 1.$$

For $l = 1$ or 2 , $g(S)$ is a maximum independent set of $B(d+2, 3)$. \square

Definition 2.14. Let S be a maximum independent set of $B(d, 3)$. We define $g'(S) \subset B(d+2, 3)$ to be the union of S , the sets $V_1(S)$, $V_2(S)$, $V_3(S)$, $V_4(S)$, $V_5(S)$ from Definition 2.12, and the sets

$$\begin{aligned} V'_6(S) &= \{yyu, u \in \{a, b\}, y \in \{d, d+1\}\}, \\ V'_7(S) &= \{uyz, y, z \in \{d, d+1\}, y \neq z, u \in \{a, b\}\}, \\ V'_8(S) &= \{(d+1)(d+1)d, dd(d+1)\}, \end{aligned}$$

which are the reverses of $V_6(S)$, $V_7(S)$, and $V_8(S)$ respectively.

Proposition 2.15. *If S is a maximum independent set of $B(d, 3)$, then $g'(S)$ is a maximum independent set of $B(d+2, 3)$.*

Proof. This proposition is proved analogously to Proposition 2.13. \square

3. ACTION OF THE SYMMETRIC GROUP

In this section, we study the interaction between \mathbb{S}_d and the four functions we defined in the previous section.

The symmetric group \mathbb{S}_d acts on the nodes of $B(d, 3)$ by $\sigma(xyz) = \sigma(x)\sigma(y)\sigma(z)$. This action preserves the graph structure, and therefore permutes the maximum independent sets. We will write $A \sim B$ to mean A and B are two sets in the same orbit under the action of \mathbb{S}_d . Notice that the functions f , f' , g , and g' are defined in a way such that if $A \sim B$, then $f(A) \sim f(B)$ and so on. Therefore, we can speak of

each of these functions as taking an \mathbb{S}_d -orbit to an \mathbb{S}_{d+1} - or \mathbb{S}_{d+2} -orbit of MISs, accordingly.

Proposition 3.1. *Let S be a maximum independent set of $B(d, 3)$. Let H , H' and H'' be the stabilizers of S , $f(S)$ and $f'(S)$, respectively. Then*

$$H = H' = H'',$$

where we identify H with its image under the inclusion $\mathbb{S}_d \hookrightarrow \mathbb{S}_{d+1}$.

Proof. We know that $H \subseteq H'$, and we must prove the other inclusion.

Let $\sigma \in H'$, and let a and possibly b be the loops of S . The set of loops must be preserved by σ and moreover, by Proposition 2.3, σ fixes each loop. We want to show that $\sigma(d) = d$. Suppose that $\sigma(d) = z \neq d$ and then $\sigma(w) = d$, for some $w \neq d$. Since w is not a loop, the node awd then belongs to the set $U_2(S)$ from Definition 2.7, and so to $f(S)$. That means that $\sigma(awd) = adz$ must be in $f(S)$. Since it begins with a , and has d in the middle, it could only be in $U_4(S)$. But $z \neq a, b$, and so $adz \notin U_4(S)$. Therefore, $\sigma(d) = d$.

Now, since $\sigma(d) = d$, σ is also an element of \mathbb{S}_d . Furthermore, it must be in the stabilizer of S . Otherwise, it should map a node of S into a node having a d . Since this is not possible, $\sigma \in H$.

The proof for H'' is completely analogous. \square

Proposition 3.2. *Let S be a maximum independent set of $B(d, 3)$. Let H , H' and H'' be the stabilizers of S , $g(S)$ and $g'(S)$, respectively. Let τ be the transposition interchanging d and $d + 1$. Then*

$$H' = H'' = \langle \tau, H \rangle,$$

where, again, we identify H with its image in \mathbb{S}_{d+2} . Notice that τ commutes with every element of H .

Proof. As in the proof of Proposition 3.1, we know that $\langle \tau, H \rangle \subseteq H'$. Now, let $\sigma \in H'$. Again, σ must preserve the set of loops in $g(S)$, and by Proposition 2.3, σ in fact fixes each loop. We will show that either σ or $\tau\sigma$ fixes d and $d + 1$. Let x, y, z and w be such that

$$x \xrightarrow{\sigma} d \xrightarrow{\sigma} y \quad \text{and} \quad z \xrightarrow{\sigma} d + 1 \xrightarrow{\sigma} w.$$

We know that $x, y, z, w \neq a, b$. Suppose that $x \neq d, d + 1$. Then we must have $dxa \in V_3(S)$ from Definition 2.12. Consider $\sigma(dxa) = yda$. This node has to be in $g(S)$, but it can only be in $V_7(S)$. That means that $y = d + 1$. Likewise, considering

$$\sigma((d + 1)za) = w(d + 1)a,$$

we have $w = d$. So $\sigma(d) = d + 1$ and $\sigma(d + 1) = d$. This contradicts our assumption about x , and implies that $x = d$ or $x = d + 1$. Analogously, $z = d + 1$ or $z = d$. That means that σ fixes d and $d + 1$ or that it transposes them. Therefore, either σ or $\tau\sigma$ is in H , and so $\sigma \in \langle \tau, H \rangle$.

The other equality follows in much the same way. \square

We now show the precise way in which our functions and \mathbb{S}_d interact.

Lemma 3.3. *Let S and S' be maximum independent sets of $B(d, 3)$. Then $f(S) \not\sim f'(S')$.*

Proof. For contradiction, suppose that there is $\sigma \in \mathbb{S}_{d+1}$ such that

$$f(S) = \sigma f'(S').$$

Let a and possibly b be the loops of S and $a' = \sigma^{-1}(a)$ and $b' = \sigma^{-1}(b)$ be the corresponding loops in S' . Let $x \neq a', b'$, $y \neq a, b$ be such that

$$x \xrightarrow{\sigma} d \xrightarrow{\sigma} y.$$

Suppose that $y \neq d$. Then the node ayd is in $U_2(S)$, and hence in $f(S)$. Therefore, $\sigma^{-1}(ayd)$ must be in $f'(S')$. But $\sigma^{-1}(ayd) = a'dx$, which cannot be in any of the sets that make up $f'(S')$. This implies that $\sigma(d) = d$. In other words, σ lies in the image of \mathbb{S}_d , and so $\sigma f'(S') = f'(\sigma S')$. However, $f(S)$ has at least one element of the form udd , and $f'(\sigma S')$ has none, so $f(S) \not\sim f'(S')$. \square

A similar result holds for g and g' :

Lemma 3.4. *Let S and S' be maximum independent sets of $B(d, 3)$. Then $g(S) \not\sim g'(S')$.*

Proof. This proof is similar to that of Lemma 3.3, but somewhat more subtle. Suppose that there is $\sigma \in \mathbb{S}_{d+2}$ such that $g(S) = \sigma g'(S')$. Let a and possibly b be the loops of S and $a' = \sigma^{-1}(a)$ and $b' = \sigma^{-1}(b)$ be the corresponding loops of S' . Let $x, z \neq a', b'$, $y, w \neq a, b$ be such that

$$x \xrightarrow{\sigma} d \xrightarrow{\sigma} y \quad \text{and} \quad z \xrightarrow{\sigma} d + 1 \xrightarrow{\sigma} w.$$

Suppose that $y \neq d, d + 1$. Then the node ayd is in $V_2(S)$, and therefore in $g(S)$. That means that $\sigma^{-1}(ayd) = a'dx$ must be in $g'(S')$. But such a node does not belong to any of the sets that make up $g'(S')$. This implies that either $\sigma(d) = d$ or $\sigma(d) = d + 1$. Analogously, we can prove that $\sigma(d + 1) = d + 1$ or $\sigma(d + 1) = d$.

Therefore, σ transposes d and $d + 1$ or leaves them fixed. By Proposition 3.2, the transposition $(d, d + 1)$ is in the stabilizer of $g'(S')$ and so by possibly multiplying σ on the right by this transposition, we can assume that σ fixes d and $d + 1$ and so it lies in \mathbb{S}_d . Therefore,

$\sigma g'(S') = g'(\sigma S')$, but $g(S)$ has at least one node of the form udd , and $g'(\sigma S')$ has none, so $g(S) \not\sim g'(S')$. \square

We now state two invariants that completely characterize maximum independent sets in de Bruijn graphs with $D = 3$. This is useful to prove that our functions f , f' , g , and g' , together with the action of \mathbb{S}_d , allow us to construct all maximum independent sets of $B(d, 3)$. In order to reverse these functions, we need the following proposition:

Proposition 3.5. *If S is a (possibly loop-less) maximum independent set of $B(d, 3)$, with loops a and possibly b . Let d' be an integer such that $a, b < d' < d$. Then,*

$$S' = S \cap B(d', 3)$$

is a maximum independent set of $B(d', 3)$ with loops a and possibly b .

Proof. Since $B(d', 3)$ is a subgraph of $B(d, 3)$, S' is clearly an independent set. Furthermore, since S has one element from each cycle except possibly a cycle that only uses the digits a and b , then S' has the same property. Therefore, S' has the cardinality of a maximum independent set. \square

Proposition 3.6. *Let S be a maximum independent set of $B(d, 3)$ with l loops, where d is at least 3. There exists a digit x such that $m_x(S) = l + 1$ if and only if there exist $\sigma \in \mathbb{S}_d$ and S' a maximum independent set of $B(d - 1, 3)$ such that $S = \sigma f(S')$ or $S = \sigma f'(S')$.*

Proof. One implication follows from the definitions of f and f' , taking $x = \sigma(d - 1)$.

Suppose now that there is such an x . We know it is not a loop by Proposition 2.3. We define the transposition $\sigma = (d - 1, x)$ and the set $S' = \sigma S \cap B(d - 1, 3)$, which is a maximum independent set of $B(d - 1, 3)$ by Proposition 3.5.

Let a and possibly b be the loops of S . We know that the node $axa \notin S$. Therefore, either xxa or axx must be in S .

Suppose that $axx \in S$. We are going to show that $S = \sigma f(S')$. To do so, we consider each of the sets that make up $\sigma f(S')$, and show that they are included in S .

The nodes of $\sigma S'$ belong to S , because of the way we defined S' .

Let us consider the nodes of $\sigma U_1(S')$. The nodes of this set are of the form xyx , xyy , yyx , xyz or yzx , for $y, z \neq a, b, x$.

- The nodes of the form xyx are all in S . Otherwise, the hypothesis cannot be satisfied.

- If $xyy \in \sigma U_1(S')$, then $ayy \in S'$. This means that $ayy \in S$, and so yyx cannot be in S . The node xyx cannot be in S either, since xyx is. So, $xyy \in S$. Analogously, if $yyx \in \sigma U_1(S')$, then $yyx \in S$.
- If $xyz \in \sigma U_1(S')$, then $ayz \in S'$ and $ayz \in S$. Since neither zxy (adjacent to xyx) nor yzx (adjacent to ayz) can be in S , xyz must be in S . The same reasoning applies to yzx .

Let us consider the nodes of $\sigma U_2(S')$. These have the form ayx . The nodes $yx a$ (adjacent to xyx) and xay (adjacent to aya) cannot be in S , which implies that $ayx \in S$. The same reasoning shows that $\sigma U_3(S') \subset S$.

Now, take a node from $\sigma U_4(S')$. That is a node of the form uxv , with u, v loops. The nodes xuv (adjacent to uxu) and uvx (adjacent to $v xv$) cannot be in S . Therefore, $uxv \in S$, and $\sigma U_4(S') \subset S$.

Finally, we know that $axx \in S$. The nodes xbx (adjacent to $bx b$) and xxb (adjacent to axx) cannot be in S . That implies that $bx x \in S$, which means $\sigma U_5(S') \subset S$.

This proves that $S \supseteq \sigma f(S')$. By cardinality, we conclude that equality holds.

If, instead of $axx \in S$ we have $xxa \in S$, an analogous procedure shows that $S = \sigma f'(S')$. \square

The following lemma is used in the proof of the next Proposition.

Lemma 3.7. *Let S be a (possibly loop-less) maximum independent set of $B(d, 3)$, with $d \geq 3$. If there exist two different digits y and z , which are not loops, such that*

$$m_y(S) = m_z(S) = l + 2,$$

then $yz y \notin S$ and $zy z \notin S$.

Proof. Suppose that $yz y \in S$. Then, by the assumptions on $m_y(S)$, there must be some $v \neq y$ such that $yvy \notin S$. Suppose that $vyy \in S$. The node zyz cannot be in S , and by the assumption on $m_z(S)$, $z v z \in S$. Therefore, the nodes zvy (adjacent to vyy), vyz (adjacent to $yz y$) and yzv (adjacent to $z v z$) are not in S . But then the cycle $zvy \xrightarrow{\theta} vyz \xrightarrow{\theta} yzv$ contributes no nodes to S , which contradicts the fact that S is maximum. If we assume that $yyv \in S$, then the cycle $yzv \xrightarrow{\theta} vzy \xrightarrow{\theta} z y v$ cannot contribute any node to S .

In conclusion, our assumption that $yz y$ is in S is inconsistent with S being a maximum independent set. By symmetry, the same holds if we assume $zy z \in S$. \square

Proposition 3.8. *Let S be a maximum independent set of $B(d, 3)$, with $d \geq 3$. There are two different digits y and z such that*

$$m_y(S) = m_z(S) = l + 2$$

and no digit x such that $m_x(S) = l + 1$, if and only if there exist $\sigma \in \mathbb{S}_d$ and S' a maximum independent set of $B(d - 2, 3)$ such that

$$S = \sigma g(S') \quad \text{or} \quad S = \sigma g'(S').$$

Proof. One implication follows from the construction of g and g' taking $y = \sigma(d - 1)$ and $z = \sigma(d - 2)$.

The proof in the other direction is analogous to the proof of Proposition 3.6. We can safely assume that $y = d - 1$ and $z = d - 2$. By Lemma 3.7, either $(d - 1)(d - 2)(d - 2)$ and $(d - 2)(d - 1)(d - 1)$ are in S , or $(d - 1)(d - 1)(d - 2)$ and $(d - 2)(d - 2)(d - 1)$ are in S . In the former case, we find that there is an S' such that $S = \sigma g(S')$. In the latter case, we find that $S = \sigma g'(S')$. \square

Corollary 3.9. *Let S and S' be maximum independent sets of $B(d - 1, 3)$ and $B(d - 2, 3)$, $d \geq 3$. Then for $\mathcal{F} = f, f'$ and $\mathcal{G} = g, g'$, we have*

$$\mathcal{F}(S) \not\sim \mathcal{G}(S').$$

Proof. This result follows from the invariants of $\mathcal{F}(S)$ and $\mathcal{G}(S')$ that are stated in Propositions 3.6 and 3.8. \square

This Corollary, together with Lemmas 3.3 and 3.4 shows that all four functions give rise to essentially different (i.e. in different \mathbb{S}_d -orbits) maximum independent sets.

4. CHARACTERIZATION OF MAXIMUM INDEPENDENT SETS

In this section, we show that the functions f , f' , g , and g' , together with the action of \mathbb{S}_d are sufficient to construct every maximum independent set of $B(d, 3)$. For the rest of this section L will denote the set of loops of S , and l will denote $|L|$. In Section 5, we will need the case when S is a loop-less maximum independent set, i.e. L is empty.

Lemma 4.1. *Let S be a (possibly loop-less) maximum independent set of $B(d, 3)$. There cannot be three different digits x , y , and z , with $x, y, z \notin L$, such that*

$$\begin{aligned} \mathcal{M}_x(S) &= \mathcal{M}_y(S) = L \cup \{x, y, z\}, \\ \mathcal{M}_z(S) &= L \cup \{x, z\} \quad \text{or} \quad L \cup \{x, y, z\}. \end{aligned}$$

Proof. Suppose that S is a maximum independent set and x , y , and z satisfy the given condition. Without loss of generality, we can assume that x , y , z , and the loops are less than 5. Then $S' = S \cap B(5, 3)$ is a maximum independent set in $B(5, 3)$ by Proposition 3.5 with $\mathcal{M}_x(S') = \mathcal{M}_x(S)$, $\mathcal{M}_y(S') = \mathcal{M}_y(S)$, and $\mathcal{M}_z(S') = \mathcal{M}_z(S)$. We can check manually that there is no such independent set S' . Therefore, there is no such independent set S . \square

Lemma 4.2. *Let S be a (possibly loop-less) maximum independent set of $B(d, 3)$. There cannot be three different digits x , y , and z , none of them loops, such that*

$$m_x(S) = m_y(S) = m_z(S) = l + 2.$$

Proof. We prove the result by contradiction. Suppose there are such x , y and z . We know that $L \cup \{x\} \subset \mathcal{M}_x(S)$ and $|\mathcal{M}_x(S)| = l + 2$. Therefore, at least one of y and z must appear between x . An analogous statement holds for y and z . Without loss of generality, suppose that y appears between x . Then xyx (adjacent to xyx) is not in S , which forces z to appear between y . That, in turn, forces x to appear between z . That is, the nodes xyx , yzx and zxy are in S . But then, none of the nodes $xyz \xrightarrow{\theta} yzx \xrightarrow{\theta} zxy$ are in S , which cannot hold. \square

Remark 4.3. Let S be a maximum independent set of $B(d, 3)$ with loops a and possibly b . There cannot be two different digits x and y such that $m_x(S) = m_y(S) = l + 1$. If there were, then xyx and yxy would have to be in S , leading to a contradiction.

Proposition 4.4. *Let S be a (possibly loop-less) maximum independent set of $B(d, 3)$, $d \geq 3$. Suppose there is no digit z such that $m_z(S) = l + 1$. Then, there must be exactly two digits x and y such that $m_x(S) = m_y(S) = l + 2$. Moreover, $\mathcal{M}_x(S) = \mathcal{M}_y(S) = L \cup \{x, y\}$.*

Proof. We just need to show that $m_x(S) = m_y(S) = l + 2$. Lemma 3.7 then implies that $\mathcal{M}_x(S) = \mathcal{M}_y(S) = L \cup \{x, y\}$.

Without loss of generality, we can assume that $m_{d-1}(S) \leq m_{d-2}(S)$ and $m_{d-2}(S) \leq m_i(S)$ for all $i < d - 2$. We know that $m_{d-1}(S) \geq l + 2$ and we want to prove that $m_{d-2}(S) = l + 2$. By Lemma 4.2, this would imply that $d - 2$ and $d - 1$ are the only digits with this property.

We prove our result by induction on d . We manually check the result for all $d \leq 5$. Now, let d be greater than 5 and consider $S' = S \cap B(d - 1, 3)$. By the inductive hypothesis, we must have one of two possibilities:

Case 1: S' has exactly one digit z with $m_z(S') = l + 1$. If $z = d - 2$, we are done.

Suppose that $z \neq d - 2$. By Remark 4.3, $m_{d-2}(S') > m_z(S')$. On the other hand, we have $m_{d-2}(S) \leq m_z(S)$, and $m_{d-2}(S')$ has to be at most $m_{d-2}(S)$. Since $m_z(S')$ is at most $m_z(S)$, we must have $m_z(S') = m_z(S) - 1$ and $m_{d-2}(S') = m_{d-2}(S)$. This means that $z(d-1)z \notin S$ and $(d-2)(d-1)(d-2) \in S$. We then have that $m_{d-2}(S) = l + 2$, as we wanted to show.

Case 2: S' has exactly two digits x and y with $m_x(S') = m_y(S') = l + 2$. We split this situation in two subcases.

Case 2.1: We suppose $x, y \neq d - 2$. By an argument similar to that of Case 1, we know that

$$\mathcal{M}_x(S) = \mathcal{M}_y(S) = L \cup \{x, y, d - 1\},$$

and that $d - 1 \notin \mathcal{M}_{d-2}(S)$. On the other hand, we know that

$$m_{d-1}(S) \leq m_x(S) = l + 3 = m_{d-2}(S).$$

By cardinality, one of x or y appears between $d - 1$ in S . Without loss of generality, suppose $(d - 1)x(d - 1) \in S$. None of the nodes $x(d - 2)(d - 1)$, $(d - 2)(d - 1)x$ and $(d - 1)x(d - 2)$ can be in S , because they are adjacent to $(d - 2)(d - 1)(d - 2)$, $(d - 1)x(d - 1)$ and $x(d - 2)x$, respectively. Thus, S cannot be a maximum independent set, a contradiction.

Case 2.2: Either x or y equals $d - 2$. Suppose $y = d - 2$. Since $m_{d-2}(S') = l + 2$, if $(d - 2)(d - 1)(d - 2) \in S$, then $m_{d-2}(S) = l + 2$, and the result follows. Therefore, we assume $(d - 2)(d - 1)(d - 2) \notin S$. This forces $x(d - 1)x \notin S$ as well.

We know that

$$(2) \quad \mathcal{M}_x(S) = \mathcal{M}_{d-2}(S) = L \cup \{x, d - 2, d - 1\}.$$

Since $m_{d-1}(S) \leq m_{d-2}(S) = l + 3$, there can be at most two digits, besides the loops and itself, which do not appear between $d - 1$ in S . Call them u and v (potentially, $u = v$.)

Case 2.2.1: Suppose $u \neq v$. We assume $u, v \neq d - 2$. That means that $(d - 1)(d - 2)(d - 1) \in S$. Since $u \neq v$, we can assume without loss of generality that $u \neq x$. Then $xux \in S$ and $(d - 2)u(d - 2) \in S$. The nodes $(d - 1)u(d - 2)$ and $(d - 2)u(d - 1)$ must be in S , because the rest of the nodes in their cycles are adjacent to something just shown to be in S . We know that $(d - 1)u(d - 1) \notin S$, because of the very definition of u . Plus, the nodes $u(d - 1)(d - 1)$ and $(d - 1)(d - 1)u$ are adjacent two one of the two nodes we just mentioned being in S . Therefore, neither of them belong to S . That gives us a contradiction.

Therefore, one of u, v must be $d - 2$, and so we have (2) and

$$\mathcal{M}_{d-1}(S) = L \cup \{u, d - 2, d - 1\}.$$

We want to show that $u = x$. Assume the contrary. Then xux and $(d-1)x(d-1)$ are in S . Therefore, by inspecting their cycles we see that both $xu(d-1)$ and $(d-1)ux$ must be in S . On the other hand, either $u(d-1)(d-1) \in S$ or $(d-1)(d-1)u \in S$.

However, $u(d-1)(d-1) \in S$ implies $xu(d-1) \notin S$, and $(d-1)(d-1)u \in S$ implies $(d-1)ux \notin S$. Therefore, $u = x$. By Lemma 4.1 applied to x , $d-1$ and $d-2$, this is a contradiction.

Case 2.2.2 $u = v$. If we assume $u \neq x$ and $u \neq d-2$ and proceed as in the previous case, we get a contradiction. Therefore, $u = x$ or $u = d-2$. In either case, Lemma 4.1 applied to x , $d-1$ and $d-2$ leads to a contradiction. \square

We now state our main result.

Theorem 4.5 (Characterization of the Maximum Independent Sets of $B(d, 3)$). *For all positive d we have:*

- (1) *Any orbit of independent sets of $B(d, 3)$ under the action of \mathbb{S}_d is obtained from the orbit of $\{000\}$ under \mathbb{S}_1 and the orbit of $\{000, 010, 111\}$ under \mathbb{S}_2 by a unique sequence of applications of f , f' , g , and g' .*
- (2) *Let S be an MIS of $B(d, 3)$. Then the stabilizer of S is generated by disjoint transpositions. In particular, this implies that the size of the stabilizer of S is a power of 2.*
- (3) *Let $b_{d,k}$ be the number of orbits of MISs in $B(d, 3)$ whose elements have stabilizers of size 2^k . Then we have the recurrence relation*

$$\begin{cases} b_{1,0} = 1, \\ b_{2,0} = 3, \\ b_{d,k} = 2b_{d-1,k} + 2b_{d-2,k-1} \quad \text{for } d \geq 3, \end{cases}$$

and the generating function

$$\sum_{d=1}^{\infty} \sum_{k=0}^{\infty} b_{d,k} t^d s^k = \frac{t + t^2}{1 - 2t - 2t^2 s}.$$

- (4) *Let a_d be the number of maximum independent sets of $B(d, 3)$. Then a_d satisfies*

$$\begin{cases} a_1 = 1, \\ a_2 = 6, \\ a_d = 2da_{d-1} + d(d-1)a_{d-2} \quad \text{for } d \geq 3, \end{cases}$$

and has exponential generating function

$$\sum_{d=1}^{\infty} \frac{a_d t^d}{d!} = \frac{t + t^2}{1 - 2t - t^2}.$$

Proof. For $d = 1$, the only maximum independent set $B(1, 3)$ consists of the unique node $\{000\}$. For the case of $d = 2$, it can be checked manually that the three orbits of maximum independent sets under \mathbb{S}_2 are the orbits of $\{000, 010, 011\}$, $\{000, 010, 110\}$, and $\{000, 010, 111\}$. Note that the first two of these are $f(\{000\})$ and $f'(\{000\})$ respectively.

Thus, the existence statement in (1) follows from Propositions 4.4, 3.6, and 3.8. The uniqueness comes from Lemmas 3.3, 3.4, and Corollary 3.9.

The statements in (2) and (3) follow from the previous result and the description of the stabilizers in Propositions 3.1 and 3.2.

Finally, the generating function in (4) is obtained by substituting $s = 1/2$ into the previous generating function, because

$$a_d = \sum_{k=0}^{\infty} \frac{d! b_{d,k}}{2^k}.$$

The recurrence follows immediately. \square

Remark 4.6. The sequence a_d is under A052608 in Sloane's Encyclopedia of Integer Sequences ([14]).

For illustrative purposes, we show the values of $b_{d,k}$, for all $d \leq 6$.

$k \backslash d$	1	2	3	4	5	6
0	1 (1,0)	3 (2,1)	6 (4,2)	12 (8,4)	24 (16,8)	48 (32,16)
1			2 (2,0)	10 (8,2)	32 (24,8)	88 (64,24)
2					4 (4,0)	28 (24,4)

In each entry, the first number between parentheses indicates the number of orbits whose elements have only one loop, whereas the second number indicates the number of orbits whose elements have two loops.

5. LOOP-LESS MAXIMUM INDEPENDENT SETS

In this section, we analyze the number of *loop-less* maximum independent sets (LMISs) of $B(d, 3)$, for all d . Recall from the introduction

that the size of an LMIS of $B(d, 3)$ is

$$\alpha^*(d, 3) = \frac{d^3 - d}{3} = \alpha(d, 3) - 1.$$

By MIS, we will continue to mean a maximum independent set *with* loops.

Definition 5.1. Let S be a maximum independent set of $B(d, 3)$, $d \geq 3$ with loops a and possibly b . We define

(3)

$$h(S) = \begin{cases} S \setminus \{aaa\} & \text{if } S \text{ has only one loop,} \\ S \setminus \{aaa, bbb, aba\} \cup \{aab, bba\} & \text{if } S \text{ has two loops } a < b, \\ S \setminus \{aaa, bbb, aba\} \cup \{baa, abb\} & \text{if } S \text{ has two loops } a > b. \end{cases}$$

Proposition 5.2. Let S be a maximum independent set of $B(d, 3)$. Then $h(S)$ is an LMIS of $B(d, 3)$.

Proof. Let S be an MIS of $B(d, 3)$. If S has only one loop, then eliminating it leaves us with an independent set of the correct size.

If S has two loops, then $h(S)$ is a set of the correct size, since the nodes we added were not already present in S . However, we must see that $h(S)$ is an independent set. Assume $a < b$. Suppose we have a node adjacent to aab . Then it is of the form abx or xaa . Since bxb and aaa are in S , then abx and xaa cannot be in S . Suppose now that we have a node adjacent to bba . Then it must be baa or xbb . Again, we know that aba and bbb are in S . Therefore, the nodes we add are not adjacent to any other nodes in the construction, and the result follows. The case $a > b$ is proved analogously. \square

Proposition 5.3. The function h is injective.

Proof. Let S and S' be two different MISs of $B(d, 3)$. Then showing that $h(S) \neq h(S')$ is just a matter of analyzing all the possible combinations of loops and their relative order in S and S' . We leave the details to the reader. \square

Lemma 5.4. Let S be a maximum independent set with two loops a and b . Let τ be the transposition of a and b . Let $S' = S \setminus \{aaa, bbb, aba\}$. Then $S' = \tau S'$.

Proof. We must show that for every node $w \in S'$ such that $a \in w$, we have $w[a \rightarrow b] \in S'$ and vice versa.

Notice that any node of S' cannot contain a and b simultaneously. The nodes that contain two a 's or two b 's are axa and bxb , and they are in S' for all $x \neq a, b$. Thus, $xay \notin S'$ for all $x, y \neq a, b$, so the nodes

that contain only one a are xya or axy for $x, y \neq a$. If $xya \in S'$, then $bxxy \notin S'$, and so xyb must be in S' in order to have one element from its cycle. We can prove that $axy \in S'$ implies $bxxy \in S'$ in a similar fashion. \square

Proposition 5.5. *The function h is surjective.*

Proof. Let S be an LMIS of $B(d, 3)$. By Proposition 4.4, we have two possibilities:

If there is a digit x such that $m_x(S) = 1$, then there is no node of the form $xxxy$ or $yyxx$. Therefore, $S' = S \cup \{xxx\}$ is an MIS of $B(d, 3)$, and $S = h(S')$.

On the other hand, if there are two digits x and y such that $m_x(S) = m_y(S) = 2$, then we have either $xxxy, yyxx \in S$ or $xyxx, xyyx \in S$. In the first case, we construct

$$S' = S \cup \{xxx, yyy, xyx\} \setminus \{xxxy, yyxx\}.$$

If $x < y$, then $S = h(S')$. If $x > y$, then $S = h(\tau S')$, where τ is the transposition of x and y . The remaining case is dealt with analogously. \square

Theorem 5.6. *Let a_d^* be the number of loop-less maximum independent sets of $B(d, 3)$. Then $a_d^* = a_d$.*

Proof. Propositions 5.3 and 5.5 show that there is a bijection between the set of MISs and LMISs of $B(d, 3)$. \square

ACKNOWLEDGEMENTS

We would like to thank Verónica Becher for introducing de Bruijn graphs to us. We also thank Alicia Dickenstein for her useful comments. We are indebted to Yukio Shibata, for kindly sharing the work of his group with us.

REFERENCES

1. Verónica Becher, *On the normality of Eulerian sequences*, Manuscript, 2009.
2. Jean-Claude Bermond and Claudine Peyrat, *De Bruijn and Kautz networks: a competitor for the hypercube?*, Proceedings of the 1st European Workshop on Hypercubes and Distributed Computers, Rennes (F. André and J-P. Verjus, eds.), North Holland, 1989, pp. 279–293.
3. CoCoATeam, *CoCoA: a system for doing Computations in Commutative Algebra*, Available at <http://cocoa.dima.unige.it>.
4. Francis Harry Compton Crick, John S. Griffith, and Leslie Eleazer Orgel, *Codes without commas*, P Natl Acad Sci USA **43** (1957), no. 5, 416–421.
5. Nicolaas Govert de Bruijn, *A combinatorial problem*, K Ned Akad Van Wet **49** (1946), 758–764.

6. Solomon Wolf Golomb, Basil Gordon, and L. R. Welch, *Comma-free codes*, Canadian J Math **10** (1958), no. 2, 202–209.
7. Solomon Wolf Golomb, Betty Tang, and Ronald Lewis Graham, *A new result on comma-free codes of even word-length*, Canadian J Math **39** (1987), no. 3, 513–526.
8. B. H. Jiggs, *Recent results in comma-free codes*, Canadian J Math **15** (1963), 178–187.
9. Yosuke Kikuchi and Yukio Shibata, *On the independent set of de Bruijn graphs*, Topis in Applied and Theoretical Mathematics and Computer Science, WSEAS Press, 2001, pp. 117–128.
10. Nicolas Lichiardopol, *Independence number of de Bruijn graphs*, Discrete Math **306** (2006), no. 12, 1145–1160.
11. Biswanath Mukherjee, *Optical communication networks*, Series on Computer Communications, McGraw-Hill, New York, 1997.
12. Pavel A. Pevzner, Haixu Tang, and Michael S. Waterman, *An eulerian path approach to DNA fragment assembly*, P Natl Acad Sci USA **98** (2001), no. 17, 9748–53.
13. Simone Severini, *Universal quantum computation with unlabelled qubits*, J Phys A-Math Gen **39** (2006), no. 26, 8507–8513.
14. Neil James Alexander Sloane,
<http://www.research.att.com/~njas/sequences/index.html?q=A052608>.
15. Edgardo Ugalde, *An alternative construction of normal numbers*, J de Th Nom Bord **12** (2000), no. 1, 165–177.

E-mail address: `dustin@math.berkeley.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, BERKELEY,
970 EVANS HALL #3840, BERKELEY, CA 94720-3840

E-mail address: `macueto@math.berkeley.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, BERKELEY,
970 EVANS HALL #3840, BERKELEY, CA 94720-3840

E-mail address: `etobis@dc.uba.ar`

DEPARTAMENTO DE MATEMÁTICA, FCEN - UNIVERSIDAD DE BUENOS AIRES,
PABELLÓN I - CIUDAD UNIVERSITARIA, C1428EGA BUENOS AIRES